DOCUMENT RESUME

ED 367 375 IR 054 923

AUTHOR Bi, Zhiwei
TITLE A Study of the Effectiveness of the OCLC CJK350 System: From the Point of View of Chinese Language Library Users.
PUB DATE Nov 90
NOTE 62p.; M.L.S. Research Paper, Kent State University.
PUB TYPE Dissertations/Theses - Masters Theses (042) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Bibliographic Databases; *Bibliographic Records; Bibliographic Utilities; *Chinese; Evaluation Methods; Foreign Students; Higher Education; Ideography; *Online Searching; Romanization; *Search Strategies; Users (Information); User Satisfaction (Information)
IDENTIFIERS *OCLC

ABSTRACT

This study examines the Online Computer Library Center's (OCLC's) CJK350 system with which one is able to input and retrieve Chinese, Japanese, and Korean bibliographic records containing both alphabetical and vernacular characters. In order to study the effectiveness of the system from the Chinese language users' point of view, a field test was conducted to collect information. Each subject was given five unique citations in Chinese scripts and asked to conduct a search in both the roman mode and CJK mode with three types of search keys: title; author or corporate author; and author and title. Their search processes and results were recorded on a search sheet. Information collected by this field test was used to evaluate the OCLC CJK Chinese part from the following aspects: whether or not the designed input methods, character subsets, and homophone qualifiers enable the various Chinese language users to have access to the OCLC Online Chinese Catalog; whether or not searches on CJK mode could be as successful as those on roman mode; and what the sample users' general perception of the OCLC CJK system are. The search sheet is included in the appendix. (Contains 12 references.) (JLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 367 375

A STUDY OF THE EFFECTIVENESS OF THE OCLC CJK350 SYSTEM: FROM THE POINT OF VIEW OF CHINESE LANGUAGE LIBRARY USERS

A Master's Research Paper submitted to the
Kent State University School of Library Science
in partial fulfillment of the requirements
for the degree Master of Library Science

By

Zhiwei BI

November, 1990

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Rosemary Du Mont

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

1205493

BEST COPY AVAILABLE

ABSTRACT

A STUDY OF THE EFFECTIVENESS OF THE OCLC CJK350 SYSTEM
FROM THE POINT OF VIEW OF CHINESE LANGUAGE LIBRARY USERS

The Online Computer Library Center (OCLC) developed the OCLC CJK350 system with which one is able to input and retrieve Chinese, Japanese, and Korean bibliographic records containing both alphabetical and vernacular characters. There are several unique features in the Chinese part of the system that are supposed to make searching of the Chinese online catalog more accurate and convenient. The diversified input methods, the choices of script form and the homophone qualifiers, enable users to search online on CJK mode when a legitimate roman citation is not available, or when roman mode search is problematic. This research, first, intends to test the various features mentioned above in order to measure, further, how effective the OCLC CJK Chinese language part is in providing diversity of searching devices for the Chinese language users with different language background. Secondly, this study also intends to test how successful searching on CJK mode could be compared to searching on roman mode. In order to do so, a field test was conducted to collect information. Twenty sample users were selected randomly from four stratified groups, the Mainland Chinese students, the Taiwan Chinese students, the Hong Kong Chinese students in the Ohio State University, and students and faculty members for whom Chinese is their second language. Each subject was given five unique citations in Chinese scripts and asked to conduct a search in both the roman mode and CJK mode with three types of search keys: title, author or corporate author, author and title. Their search processes and results were recorded on search sheet. Information collected by this field test are used to evaluate the OCLC CJK Chinese part from the following aspects:

1) whether or not the designed input methods, character subsets and homophone qualifiers enable the various Chinese language users to have access to the OCLC Online Chinese Catalog;

2) whether or not searches on CJK mode could be as successful as those on roman mode;

3) what the sample users' general perception of the OCLC CJK system are.

Master's Research Paper by

Zhiwei Bi

B.A., University of California, Irvine, 1988

M.L.S., Kent State University, 1990

Approved by

Adviser Ann Allen Date November 28, 1990

TABLE OF CONTENTS

	PAGE
I. Introduction	1
1. The Arrival of the CJK Systems	1
2. The Users of the OCLC CJK System	2
3. The Purpose of the Study	3
II. Problems in Arrangement of Chinese Material and the Literature Review	4
1. Problems in Management of Chinese Language Material	4
a. The Ideographic Chinese Characters	5
b. Monosyllables and Homophones	5
2. The Need for CJK System	7
3. A Literature Review	8
III. OCLC CJK350 System	9
1. General Functions and Components of the OCLC CJK350 Workstation	9
2. Forms of CJK Characters	10
3. Methods of Input and Character Subsets	11
4. Search Keys	12
1) Numeric Search Keys	12
2) Derived Search Keys in Roman Alphabet	12
3) Derived Search Keys in CJK Character	14
5. Searching on CJK350	15
IV. The Chinese Part of the CJK350	16
1. Wade-Giles and Pinyin	16
2. Full and Simplified Forms	18
3. Ts'ang-chieh and The CJK Input Code Dictionary	19
4. Reduction of Homophones	21
V. The Objectives of the Research	23
VI. The Methodology	24
1. Samples	25
2. Procedure of the Field Test	26
3. Instrument	27
4. Notes	27
VII. The Findings	28
1. Search Pattern	29
2. Sample Subjects' Perception of the OCLC CJK System	32

3. Search Results _____ 35
1) Distribution of Records _____ 37
2) Problems in Roman Search Keys _____ 38
3) CJK Search Superior to Roman _____ 41
4) Yes or No Records _____ 42
5) E Records _____ 43

VIII.The Conclusion _____ 44

Bibliography

Appendix A: Search Sheet

LIST OF ILLUSTRATIONS

	PAGE
Figure 1. Sample of OCLC CJK Record	48
Figure 2. Sample of OCLC CJK350 Card	49
Figure 3. CJK Message Blocks in the Status Line	50
Figure 4. Graphic Elements Used in Ts'ang-chieh Entry	51
Figure 5. Pinyin, Mandarian Phonetic Symbols, and Wade-Giles Table	52

List of Tables

	Page
Table 1. Search Pattern	34
Table 2. Perception A	35
Table 3. Perception B	35
Table 4. Distribution of the Six Types of Records	41
Table 5. Comparison of Found and Not-found Rates in CJK and Roman Searches	41

I. INTRODUCTION

During the 1970s, many major libraries in the United States went online to share a common bibliographic database with others in support of their cataloging, circulation, acquisition, interlibrary loan, and other library functions while Chinese and other East Asian bibliographic records were still excluded from the national and local databases. Chinese collection operations and services were isolated from the mainstream of library automation. One of the major causes for this exclusion was due to the fact that computers at that time were not able to handle the storing and retrieving of the Chinese ideographic characters.

1. The Arrivals of the CJK Systems

Until the late 1979, RLIN (the Research Libraries Information Network of the Research Libraries Group,) by introducing the computer technology for automation of Chinese, Japanese and Korean (CJK) vernacular scripts, started the project of getting East Asian bibliographic records into national database. On September 12, 1983, LC entered the first CJK vernacular record into the RLIN database [Tucker 1982]. The completion of the RLIN CJK system marked the beginning of online operations of East Asian libraries in the States. In the same year, OCLC (the Online Computer Library Center), launched its own CJK program and completed its development in 1986 [Wang 1986]. The automation of East Asian languages has been brought to a more sophisticated stage. The success of the both systems has made East Asian

cooperative cataloging and resources sharing a true reality.

2. The Users of the OCLC CJK System

As Andrew Wang stated in his article " The OCLC Library Network of Chinese, Japanese and Korean Characters ", " the purposes of the OCLC CJK 350 system are to reduce libraries' cost of processing materials in CJK languages, to reduce the time required to process these materials and to increase the availability of bibliographic and location information of materials in CJK languages to librarians, scholars, and students world-wide." [Wang 1988]. Being able to store and to retrieve East Asian language records, the OCLC CJK system serves its dedicated users such as catalogers and other librarians in several ways. It can be used as a cataloging workstation. Catalogers use it to input cataloging information, to find out whether the material has been cataloged by any other member libraries, to decide whether the cataloging can be done by copying the OCLC record. It can also serve as an East Asian material online union catalog. Bibliographers use it to verify bibliographic records, to check for location information and other related library activities.

Thanks to its unique function as an online union catalog, the OCLC CJK system can be also beneficial to another group of users. They are the casual users such as the East Asian languages library patrons. These library patrons could probably consult the

OCLC CJK database as they use an online public library catalog for their bibliographic requirement and location information.

3. The Purpose of the Study

Since its creation in 1986, the OCLC CJK system has been widely used, but almost exclusively by librarians in at least 30 libraries; only a few library patrons, however, have direct access to the system. In fact, very few users of the library, whether speakers of an Asian language or students of those languages, have known about the system. The OCLC CJK Workstation so far remains a librarian's tool rather than a online East Asian language material catalog for public users.

It is the purpose of this field test, first, to the bring the Chinese language users' attention to the OCLC CJK system; secondly, to test how effectively the system could provide casual users with efficient input methods to access the OCLC Chinese online catalog; thirdly, to investigate the degree of success or failure in searching on the Chinese part of the OCLC CJK system. By doing so, the study may be able to generate a preliminary and primitive evaluation of the OCLC CJK Chinese part from Chinese language library users' point of view. This information will be helpful to librarians in understanding how effective the system could be as a public workstation, in determining whether the new terminal for public use is worth the purchase, and in considering what elements should be included in user training when the system

become a public workstation.

The next chapter will present the common problems in arranging the Chinese materials. It will also discuss the previous studies on the CJK systems. In Chapter III, the general structure of the OCLC CJK system will be described. In Chapter IV, special features of the Chinese language part of the CJK such as input methods, character subsets and homophone reducers will be discussed in detail. Chapter V will present the objectives of this study, and Chapter VI deals with the experimental methodology of the study. Finally, the findings and the survey analysis will be presented.

II. Problems in Arrangement of Chinese Materials and the Literature Review

1. Problems in management of Chinese language material

In order to understand how the special functions of the Chinese language part of the OCLC CJK350 were designed to overcome the difficulties in handling Chinese language material, problems in arranging the Chinese language material have to be addressed. In fact, the management of Chinese materials in the United State libraries had been problematic for almost one century since the first real attempt to adopt a cataloging and classification scheme in the 1910s [Kwei 1931, 13]. These problems are due to the unique nature of the language.

a. The Ideographic Chinese Characters

During the period of 1911 to 1957, Chinese materials were treated with special classification and cataloging schemes and segregated from the main stream of cataloging because of the unique ideographic characteristic of the language. The card catalog was filed according to the form of the Chinese characters, such as radical (which means the principal part of a character which is common to many similar character), and strokes (parts that consist of a character), etc. (In both cases, all the characters are arranged by stroke number of either the radical or the whole character. Those possessing a lesser number are followed by those having more.) During this period, each Chinese collection had its own way in treating its materials; there was no standard system. This situation made the sharing of cataloging and resources very difficult.

b. Monosyllables and Homophones

In 1958, Library Congress began to catalog and classify Chinese and other East Asian langauge materials in the same manner as their Western collection. This was the first time romanization of the Chinese character was used for entry in cataloging. Romanization means simply taking the sounds of a non-alphabetic language and rendering them into a Latin script so as to be more accessible to the Westerner. The romanization system that has generally been used in cataloging Chinese materials in the United States is the Wade-Giles system, and these phonetic

scripts represent the sound of Beijing dialect or mandarin.

Compared with the former special treatments for the Chinese material, this evolution by LC is advanced in that it allows Chinese bibliographic records to be integrated into the Western language bibliographic records. However, the system of adapting western library techniques to Chinese cataloging is far from perfect. One of the main defects is that, in many case, romanizations of the Chinese character are very difficult for readers to understand. This is due to another nature of the Chinese language. The Chinese language is monosyllabic and a large number of characters under the same sound have different meanings. For example, the phonetic script "Zhong" bears the same sound for at least five characters that have different meanings. (Zhong: 中、钟、终、宗、忠 .) Further more, The Chinese language is a tone language, for which a large number of characters that bear the same romanization have different tones; thus, they have different meanings. For instance, Ma, Ma, Ma, Ma, according to their four different tones could be the following four characters: 妈、麻、马、骂 . A phrase or a sentence like this, "Ma ma ma," could possibly have three meanings when it is pronounced in different tones: 1) the mother of the little horse; 2) to accuse the mother; 3) the mother accuses. As a result, it would be very confusing to use a catalog without the presence of the actual Chinese characters. Thus, many libraries that adapted western library techniques to Chinese material still had to keep

their old card catalog while some other libraries just ignored what the LC had done and maintained their old card system file by the stroke count.

2. The need for CJK system

When western language materials were put online in 1970s, the special nature of the Chinese language prevented Chinese materials from being automated. Because of that, computers designed for dealing with roman languages are incapable of storing and retrieving ideographic characters: Chinese, as well as Japanese and Korean bibliographic records were again excluded from national and local databases. The fact that romanization would not work perfectly for the Chinese materials and the technology was not able to handle the ideographic characters, caused great concern among East Asian librarians and users, supporting funding organizations and the major bibliographic utilities. These concerns and great efforts by specialists have brought about the completion of both the RLIN and OCLC CJK Systems. Now East Asian language materials can be stored and retrieved at library computer terminals as readily as Western language publications. Since 1986, at least thirty East Asian libraries have joint the OCLC CJK systems and benefited from the wonders of this new technology, and the ability to input and retrieve Chinese, Japanese, and Korean bibliographic records containing both vernacular characters and their romanizations. The presence of the Chinese vernacular character in bibliographic

records sweeps away the confusion that usually exist in the romanized records.

3. A Literature review

A review of the previous publications on the CJK programs shows that attention has been directed to the comparison of the overall advantages and disadvantages of the RLIN and OCLC CJK systems. Very few studies have been conducted to measure the effectiveness of the systems in providing access to the CJK database for users, especially for the casual users such as Asian Studies students and faculty members. In 1984, a survey of current status and future trends of East Asian Library automation in North America showed that in only eight out of seventeen RLIN libraries, did the public have access to the CJK terminals. Based on the telephone conversations that I had with three workers involved in OCLC CJK program, the OCLC CJK system so far is basically used as a cataloging workstation by libraries. However, it is important for us not to neglect its other function, that of an online union catalog for East Asian materials, which has been needed for many years. The system can be beneficial to public users to locate the East Asian language materials they need nationwide.

Regarding the OCLC CJK350 itself, problems do exist in actual searching on the system. In 1988, a student of the Kent State University School of Library Science, Jeong Hyun Kim, conducted a

study on the search failure of OCLC CJK350 system. The finding suggests that users may confront difficulty in a Korean search, since the parallel use of Hangul (unique Korean characters) and Hancha (Chinese characters) was not considered in the design of the CJK350.

III. OCLC CJK350 SYSTEM

1. General Function and the Components of the OCLC CJK350 Workstation.

The OCLC CJK terminal is called a CJK350 Workstation. It is an enhanced M300 workstation based on the IBM PC/XT Configuration. Its keyboard is about the same size as a regular typewriter with additional function and control keys. It is a phonetic or coding entry system rather than a character-component entry system [Wei 1989]. By using the RLIN East Asian Character Code (REACC), it can produce about 16,000 characters.

There are three software packages for the OCLC CJK system: 1) CJK Online Cataloging Package; 2) CJK Card Production Package; 3) CJK Word Processing Package. The first is essential for online cataloging of bibliographic records containing CJK vernaculars. (See Figure 1. for an example of Chinese bibliographic records containing both the romanized form and vernacular script.) The card production software allows cards in CJK characters to be printed by a local printer from online records. CJK characters

are printed in a 24 x 24 dot matrix. (See Figure 2 for a sample of CJK card). The word processing package combines Chinese, Japanese, and Korean language word processors into one package used for word processing and file management. It is not immediately associated with library operations.

In general, the OCLC CJK system possesses three basic functions: 1) it is able to process East Asian language bibliographic records with vernacular characters; 2) it makes bibliographic search possible on both roman mode and CJK mode with various input methods; 3) it allows a complete set of catalog card in CJK character to be printed in the user's library.

2. Forms of CJK Characters

OCLC CJK350 system can generate seven forms of CJK characters, two for Chinese, three for Japanese, and two for Korean. They are:

1) Chinese
   a. full character (CF)
   b. simplified character (CS)
2) Japanese
   a. kanji (JJ)
   b. katakana (JK)
   c. hiragana (JH)
3) Korean

a. hancha (KC)

b. hangul (KH)

3. Methods of Input and Character subsets

The OCLC CJK350 system provides five input methods as they are listed below:

1) Modified Hepburn (HP);

2) McCune-Reischauer (MR);

3) Wade-Giles (WG);

4) Pinyin (PY);

5) Ts'ang-chieh (TC).

The first two methods, Modified Hepburn and McCune-Reischauer, are used for the input of Japanese and Korean characters. Number 3 and 4, Wade-Giles and Pinyin are two pronunciation-based input methods used for Chinese. Number 5, Ts'ang-chieh, the only character-based method, applies to all three languages.

The following chart summaries the input methods, and the characters they generate. [Wang 1988]

	CHINESE		JAPANESE			KOREAN	
	Full	Simplified	Kanji	Katakana	Hiragana	Hancha	Hangul
Ts'ang-chieh	X	X	X			X	
Pinyin	X	X					
Wade-Giles	X	X					

Modified Hepburn	X	X	X		
McCune-Reischauer				X	X

4. Search Keys

The OCLC CJK350 provides three categories of search key 1) numeric, 2) roman alphabet, 3) CJK character. They are explained as the following:

1) Numeric search keys,

a. LCCN: the Library of Congress Card Number.

b. ISBN: the International Standard Book Number.

c. ISSN: the International Standard Serial Number.

d. Coden: a five-letter code assigned to serials by Chemical Abstract Service.

e. OCLC Control Number: a unique number assigned by the OCLC Online System to each bibliographic record as it enters the OCLC Online Union Catalog.

f. Government Document Number.

g. Music Publisher Number: plated and publishers' numbers for printed music, and serial numbers and matrix numbers for sound recordings.

2) Derived search keys in roman alphabet

a. Title search key (3,2,2,1) consists of the first three letters of the first word in the title, excluding an initial article, followed by a comma, the first one or two litters of the second wording the title, another comma,

the first one or two letters of the third word in the title , another comma, and the first letter of the fourth word in the title.

b. Personal name search key (4,3,1) consists of the first four letters of the author's surname, followed by a comma, the first one, two, or three letters of the author's forename, another comma, the author's middle initial which is optional.

c. Corporate name search key (=4,3,1) consists of an equal sign, followed by the first four letters of the first significant word in the name, a comma, the first one, two, or three letters of the word following the first significant word, another comma, and the first letter of the next word which is optional.

d. Name/Title search key (4,4) consists of the first three or four letters of the first word in the author's surname, corporate name, or uniform title, followed by a comma, and the first three or four letters of the first word in the title excluding initial articles.

These eleven search keys listed above are all applicable for retrieving roman alphabet records in the OCLC Online Union Catalog as well as CJK records. In the OCLC Online Union Catalog, each field that contains CJK characters links with another field that contains corresponding information in its romanized form. The search must be conducted on the alphabet mode, and only

certain romanization systems are legitimate input methods. For example, in order to search bibliographic records with Chinese characters Wade-Giles has to be used.

3) Derived search keys in CJK characters

In addition to the above, there is the third kind of unique search key - search keys in CJK characters:

a. Title search key (ti:5) begins with a prefix "ti:", followed by the initial one, two, three, four, or five CJK characters in the title. The more characters there are in the search key, the fewer records it will retrieve;

b. Personal name search key (pn:4) begins with a prefix "pn:", followed by the initial one, two, three, or four characters in the corporate name;

c. Corporate name search key (cn:) begins with a prefix "cn:", followed by the initial one, two, three or four characters in the corporate name;

d. Name/Title search key (nt: 4, 4) begins with a prefix "nt:), followed by none, or the first character in the name, a comma, and the initial one, two, three, or four characters in the title [OCLC 1986].

These four search keys in CJK characters will only retrieve records with CJK characters; when one search key retrieves more than one record, these records are sorted by the romanized alphabet instead of the CJK vernacular. The prefix part of the search key must be inputted on the alphabet mode while the rest

of it on CJK mode. The CJK character search key can be qualified by type of material, year of publication, etc., as the numeric and alphabet search keys can be, and all qualifiers must also be conducted on roman mode.

5. Searching on CJK350

There are generally seven steps to use CJK character search key in searching CJK character records.

Step 1. Choose and key in search key's prefix, (ti:, pn:, cn:, nt:), in the regular screen, the alphabet mode.

Step 2. Press <Ctrl> and <F11> to switch to CJK or vernacular mode from roman, the CJK input status line will appear at the bottom of the screen. (The CJK input status line have five message blocks. See Figure 3.)

Step 3. Press <Ctrl> and <F9>, then type in the selected input method from TC, PY, WG, MH, MR. The input method will appear in the first message block of the CJK input status.

Step 4. Press <Ctrl> and <F10> to select the character form, from FC, SC, JC, JK, JH, KC, KH; it will appear in the second block of the status line.

Step 5. Key in the character-part of the search key in the third block of the line.

Step 6. Press <Ctrl> and <F11> to switch back to roman mode,

then key in qualifier or qualifiers.

Step 7. Press <DISP SEN> to retrieve CJK records.

IV. The Chinese Part of the CJK350

Although the OCLC CJK system lacks a subject search capacity, it has been well used by libraries for cataloging and other activities. Beside of its unique feature of processing bibliographic records in both English and East Asian vernacular characters, its various phonetic input methods based on reading instead of on components of the CJK character may be also applicable for Chinese language users with different language background in bibliographic searching. In order to make this point clear, some changes in the Chinese language, its romanization system and writing system, happened in Mainland China and other Chinese-speaking areas, and problems brought with these changes should be addressed.

1. Wade-Giles and Pinyin

It was mentioned earlier that Wade-Giles is the romanization system used in cataloging Chinese material in the United States since the 50s. However, in the mid-fifties, with the original intent of eventually doing away with Chinese characters altogether, linguists in the People's Republic of China decided to develop a new and more consistent system of romanization; it was officially adopted in 1958 and is currently the official

system of romanization in the PRC and Hong Kong where most Chinese library materials come from. This system is Pinyin. When relations with the West opened up in the late seventies it became apparent there was a conflict between these two romanization systems.

Wade-Giles and Pinyin are two quite different systems. For example, the name of a famous Chinese beer, in Wade-Giles, is Tsing-tao, in Pinyin it is rendered Qingdao. Another example is the name of one of the input method of CJK350, Ts'ang-Chieh, in Pinyin it should Cangjie. When books purchased from mainland China and Hong Kong arrive, they are cataloged in Wade-Giles system, even though this rendering may be at odds with the official romanization. This inconveniences not only the cataloger but also the user. For instance, if the titles of works referred to in a particular reference work are in Pinyin, a user has to translate from Pinyin into Wade-Giles before he can search it in the catalog. Today's user may well not be familiar with Wade-Giles for Pinyin is now the romanization system preference (as a pedagogic tool in helping students in pronunciation of Chinese) for most Chinese language programs in the United States. The availability of the two input methods, Wade-Giles and Pinyin, provide a potential solution for this inconvenience. It is supposed that one who does not know Wade-Giles will be able to do his search in Pinyin or vice versa. It is one of this study's intent to see how effective a solution this is in practice.

2. Full and Simplified Forms

As was shown in an earlier section, the three methods of inputting Chinese characters can generate not only the full Chinese character but also the simplified form. It should be noticed that the Chinese script has been changing throughout its history. In Qin and Han dynasties it underwent large-scale revision; in the early twentieth century, the reform and simplification of the traditional script was considered as an important part of the general reform movement. In 1956, the government of the People's Republic of China issued a list of 515 simplified characters. In 1964 a further list of more than 2,000 simplified characters was issued; many of them resulting from the simplification of common radicals and phonetic components was put into effect (Norman 1988). These simplified characters have been in widespread use in Mainland China, while in other Chinese-speaking areas in the world such as Hong Kong and Taiwan, the traditional script or full characters are still being used.

In the practice of cataloging, the form of the characters of a book will be recorded according to the title page. For a book from Mainland China having a simplified character title, its bibliographic record will appear in simplified form character. On the contrary, for a book from Hong Kong or Taiwan, its record will appear in the complicated form. However, one can use either form to retrieve the wanted record.

The fact that the OCLC CJK system Chinese part provides two kind of phonetic input methods and the fact that it can generate both simplified and full character could possibly make bibliographic searching more convenient for Chinese language users. The Pinyin romanization system and the simplified character may be applicable for users from Mainland China while Wade-Giles and full characters maybe suitable to users from Taiwan and other areas where Pinyin and simplified characters are not used.

3. Ts'ang-chieh and The CJK Input Code Dictionary

In addition to the various phonetic input methods, the OCLC CJK system also provides one character-based input methods, Ts'ang-Chieh. It can be used for generating Chinese characters in both full and simplified forms, as well as the Japanese kanji and the Korean hancha derived from the Chinese. As was mentioned earlier, the two most popular romanization systems, Pinyin and Wade-Giles, are the phonetic scripts representing the sound of Mandarin. For a Chinese language user who does not speak Mandarin, who has neither the knowledge of Wade-Giles nor Pinyin, the two phonetic input methods provided by the OCLC CJK system are not applicable; thus, Ts'ang-chieh may become the only useful input method for them.

However, one major problem with this method is the complexity of the alphabet codes for the Chinese character. In order to use

it as an input method in the searching process, one has to be familiar with the alphabet codes unique for each character. One tool that could be helpful is The CJK Input Code Dictionary. It serves as a ready reference hand book for composing input codes; it also prόvides cross-references among the romanization systems).

These four volumes are:

Volume 1: Chinese Language

Part 1

Section 1: Tsang-Chieh -- Wade-Giles

Section 2: Wade-Giles -- Tsang-chieh

Section 3: Wade-Giles -- Pinyin

Volume 1: Chinese Language

Part 2

Section 1: Tsang-Chieh -- Pinyin

Section 2: Pinyin -- Tsang-Chieh

Section 3: Pinyin -- Wade-Giles

Volume 2: Japanese Language

Section 1: Tsang-Chieh -- Modified Hepburn

Section 2: Modified Hepburn -- Tsang-Chieh

Section 3: Modified Hepburn for Hiragana and Katakana

Volume 3: Korean Language

Section 1: Ts'ang-chieh -- McCune-Reischauer

Section 2: McCune-Reischauer -- Ts'ang-chieh

Section 3: McCune-Reischarer for Hangul

In both parts of the Volume 1. the input codes are arranged alphabetically except for the Ts'ang-chieh system. As the title of each section of the dictionary indicates, one is able to use one input code system to chase the others. For example, One can find the Pinyin romanization for the Wade-Giles "ts'ang" as "cang", and "chieh" as "ji" by looking in Section 3 of Part 2, under "t" and "c" sections. One can also find the Ts'ang-chieh input code for the Pinyin romanization "ma" as "SQSF" by looking under "ma" in Section 2 of Part 2. It could be still difficult, however, to locate a particular input code of Ts'ang-chieh because the input codes of Ts'ang-chieh are arranged by graphic elements assigned to English letters (See Figure 3). Without some intensive training it could be difficult for one to determine under which graphic element a character is listed. So far, this input method is primarily used by dedicated users, such as catalogers. Questions then arise.. How successful it could be for a casual user, who does not have intensive training, in conducting a search by using the dictionary? What kind of problems he might come across and what kind of aid should be prepared for these users?

4. Reduction of Homophones

As has been discussed in the introduction, the homophone is one

of the problems in arranging Chinese material; so is it in retrieval of Chinese characters in computer systems. For example, when one keys in the romanization of the Chinese character , "zhong", the OCLC CJK Workstation generates 32 homophones of the character. The system seems to improve the situation by providing two optional homophone reducers. The first of them is the application of tone marks. The Chinese language is a tone language; tones are used to differentiate meaning. A user of Wade-Giles and Pinyin may add to the end of input keystrokes a numeral 1, 2, 3, 4 or 5 to signify the tone of the character desired. In the example above, if "zhong" was inputted as "zhong 3", the numbers of homophones would be reduced from 32 to 8; search would be more effective and less time-consuming. One can also use the first letter of Ts'ang-Chieh input code of the character desired to further eliminate homophones. The application of this is simply spelling out the radical of the character, thus eliminating other homophones that do not share the same radical. In the case of the last example, the use of "zhong 3 H" will reduce the numbers of the homophone from 32 to 3.

In an actual search, if one does not use the homophone qualifiers, when homophones occur, the OCLC CJK350 workstation will beep and display up to eight characters at a time in the homophone block of the status line (the fourth block of the CJK message status line. See Figure 3). Each character so displayed

subsets, full and simplified characters to search on the system?

b. Is The CJK Input Code Dictionary helpful and easy to use according to the Ts'ang-chieh users? How successful a search could a casual user (without intensive training) conduct by consulting the dictionary? What kinds of problems might he come across and what kinds of aids should be prepared for these users?

c. Are the two homophone qualifiers, tone and radical, effective in reducing homophone according to their users?

The second aspect of this study is concerned with the quality of the search result on CJK mode. Two questions need to be answered:

a. What are the hit rate on both the roman and CJK mode searches with same kind of search key?

b. When will a CJK mode search be more effective in retrieving Chinese bibliographic records, and when will it not?

In addition to these two aspects, the study is also concerned with sample users' perceptions of the OCLC CJK system. Information, such as whether the system is user-friendly; whether sample user would like to use it again, will be collected.

VI. Methodology

In order to collect information required for evaluating the Chinese language part of the OCLC CJK350 system from the Chinese

language users, a field test on the system was conducted.

1. Sample users

Twenty sample subjects from two categories, users who are native Chinese speakers and users for whom Chinese is their second language, were invited to participate in the field test. They were selected randomly from four stratified groups and each group consists of five subjects:

1) Group ML - subjects chosen from a list of the Mainland Chinese students of the Ohio State University;

2) Group TW - Taiwan Chinese students of the OSU;

3) Group HK - Hong Kong Chinese students of the OSU;

4) Group CSL - Chinese studies students and faculty members for whom Chinese is a second language.

These four groups of Chinese language users have different language backgrounds. Most of the Mainland Chinese students are able to understand Mandarin and familiar with Pinyin and simplified Chinese characters. The Taiwan students also have the knowledge of Mandarin, but they do not usually know Pinyin. The romanization system used in Taiwai is called "Mandarin Phonetic Symbols" [Liang 1972]. In fact, Pinyin and Mandarin Phonetic Symbols are one system in two forms; Pinyin is the use of the Latin alphabet to indicate the pronunciation of the Chinese characters while Mandarin Phonetic Symbols is the use of syllabic script. (See Figure 5 for a sample of Pinyin and Mandarin

Phonetic Symbols Table.) Many Hong Kong Chinese students do not speak Mandarin but Cantonese. As the Taiwanese students do, they also use complicated form of Chinese characters. The non-native Chinese speakers in the fourth group would be familiar with various romanization systems and character forms depending on their learning experience. With particular language background, these Chinese users should need a proper search device in order to search on the system successfully. It was supposed that an analysis of the search patterns, and success and failure of these four sample user groups, should suggest to the researcher whether the OCLC CJK system is effective in providing various Chinese language users with sufficient input devices in their bibliographical searches.

2. The Procedure of the Field Test

All together, 100 titles were randomly selected from the new Chinese Card Catalog in the Ohio State University Main Library. Each sample subject was given five unique citations written in Chinese scripts, and asked to conduct searches on the OCLC CJK350 Workstation located on the third floor of the OSU Main Library. Each item was searched on both the CJK and roman modes with three types of search keys. For the CJK mode search, search keys used are title (ti:5), personal name (pn:4) or corporate name (cn:4), and name /title (nt:1,4); for the roman mode search, they are author (4,3,1) or (=4,3,1), title (3,2,2,1), and author and title (4,4).

BEST COPY AVAILABLE

The researcher was with the sample users during each search in order to provide necessary technical instruction, to discover difficulties confronting the sample subjects, and to record the process of each search.

3. Instrument

In order to collect required data from the field test, a "Search Sheet" was formed. The search sheet consists two parts:

Part I. The Searching form

The Searching form includes the following information

a. input methods used;

b. script forms used;

c. input character codes;

d. homophone qualifiers;

e. result (with printouts)

Part II. Eight questions about the subject's language background and their reaction and perception of using the OCLC CJK Chinese part. (For a sample of the Search Form See Appendix A)

4. Notes

a. Use of the two qualifiers of type of material and year of publication provided by the system were not allowed in the field test searching on either mode, considering that usage of them would complicate and confuse the comparison hit-rate count.

b. In order to conduct the survey, the following equipment and materials were used:

1) the OCLC CJK350 Workstation located on the third floor of the Ohio State University Main Library;

2) a four-volume set of The CJK Input Code Dictionary.

3) a table of Pinyin and Mandarin Phonetic Symbols.

c. The method used for data analysis in this study, will be mainly descriptive statistics. Tables and charts will be used to provide more informative data, and the task of the data analysis will be done by the researcher.

d. Because it is assumed that the participants of the field test are already familiar with the English-language key board, and that they are capable of reading English instruction when doing the search, library users who neither know how to use the key board nor read English are not included in the study. Thus, their perception of the system is excluded.

VII. Findings

Eighteen sample subjects out of a total of twenty (90%), actually participated in the field test. Among them were fifteen members of the three principal groups Mainland (ML), Taiwan (TW), and Chinese as Second Langauge (CSL), and three from the Hong Kong (HK) group. The following sections will describe the pattern

of search by these groups of Chinese language users and their perceptions of using the OCLC CJK system. This will be followed by an analysis of their search results.

1. Searching Pattern

Table 1 shows that 89% of the participants used Pinyin input method when searching in CJK mode. It is to be expected that sample users of the Mainland Chinese students (ML) would use Pinyin as Pinyin has been the standard romanization system used in the PRC for years. It is interesting to see that all the Taiwan participants (TW), as well as those for whom Chinese is a second language (CSL), also chose to use Pinyin despite the differences in their language background. (The TW students were not familiar with either Pinyin or Wade-Giles, while all members of the CSL group had knowledge of both.) This searching pattern suggests that Taiwan sample users can perform searches on the OCLC CJK terminal using Pinyin input method if a Pinyin and Mandarin Phonetic Symbols Table is provided.

The three groups discussed above (ML, TW, CSL) are native speakers or students of Mandarin Chinese. Knowledge of Mandarin enables them to utilize a phonetic input method such as Pinyin, a romanization developed for the Mandarin dialect, to perform their search. However, there is another group of Chinese language users, the non-Mandarin speakers, for whom this system is of little use. The examples of this group are the two Hong Kong

participants. Neither was successful in using either of the two phonetic input methods, but they were able to make use of Ts'ang-chieh, the character-based input methods of the OCLC CJK system. The success of these two Hong Kong students in completing their searches on the OCLC CJK suggests that Ts'ang-chieh input method can be helpful in providing access to the OCLC Chinese online catalog for non-Mandarin speakers.

Table 1 also indicates that Wade-Giles (WG) input method was not as favorable as Pinyin (PY). There were two Taiwanese sample subjects trying to use the method at the beginning of their searches. However, both of them later shifted to Pinyin when they discovered the Wade-Giles romanization to be rather difficult to learn during the search. One CSL member also started with Wade-Giles, but continued his search by using Pinyin after a try at it. He found Pinyin easier because this input method does not require as many diacritic as does Wade-Giles. Because of this, searching with Pinyin was more convenient and less time-consuming, especially for those who were not very familiar with the search procedure. (Figures for the use of Wade-Giles are presented in parentheses in Table 1.)

Another feature of the OCLC CJK system that this study intends to investigate is the usefulness of the two forms of Chinese characters, simplified and full. Table 1 shows that all of the ML and two of the CSL sample users, 39% of all the participants,

chose the simplified form (CS); the remaining eleven subjects (from TW, HK and CSL groups), who constituted 61% of the total, used full form (CF). Results of their searches indicated that CJK records can be retrieved by inputting either form of the characters, although in certain records characters might be full or simplified. In other words, user were able to retrieve a record in full form even though his/her input was in simplified, or vice versa. The fact that both the Mainland Chinese participants, who are more familiar with simplified characters, and the Taiwanese, who generally have knowledge of only full characters, succeeded in completing their searches indicates that the availability of two forms of Chinese characters can make OCLC Chinese online catalog more accessible to Chinese language users familiar with only one of the two forms.

Another element considered in the study is the efficiency of the homophone reducers. Table 1 shows that fourteen of a total of sixteen sample subjects (87%) that chose the Pinyin phonetic input method used the tone homophone reducers. Only two (13%) did not make use it as they were unable to distinguish the tone of the characters they were searching. Table 2 presents these 14 users' opinions on the efficiency of this feature and it indicates that ten out of a total of fourteen users (71%) found it helpful; the remaining four (29%) felt it very impressive.

None of the participants used the radical homophone reducer

during the field test. This is simply because the application of a radical homophone reducer requires familiarity with the Ts'ang-chieh (TC) input codes. Unless previously trained in these codes, anyone wanting to use them must consult The CJK Input Code Dictionary and to do so would clearly slow down the searching process. Moreover, homophone reducers are not necessary for the Ts'ang Chieh system because almost every character has its unique code. An analysis of the sample users' perception and their search pattern indicates that tone qualifier is useful in phonetic method searching, and radical qualifier is not appropriate for casual users who lack intensive training in the Chinese character codes.

2. Sample Subjects' Perception of the OCLC CJK System

Sample subjects' perceptions of the system focus on four aspects: a) the efficiency of homophone reducers (which was addressed in the two previous paragraphs); b) the convenience of The CJK Input Code Dictionary; c)the search itself; d) and the willingness of these subjects to use the system again in the future. Information on these four aspects are collected in Table 2. Perception A and Table 3. Perception B., and it can be described as follows.

There were only two out of a total of eighteen sample users who selected Ts'ang-chieh as an input method. Table 2 shows of these two, one considered the CJK Input Code Dictionary easy to use,

the other found it rather difficult. Because of the absence of the other two Hong Kong sample subjects, opinions on Ts'ang-chieh input method and the dictionary are not as well represented as one should expected from this study. Further studies are necessary to test the efficiency and convenience of this method.

Table 3 indicates the sample subjects' perceptions of searching on the OCLC CJK system. Seventeen percent (three out of a total of eighteen) of sample users expressed the opinion that the searching was very difficult; 44% (eight) considered it not very difficult; 28% (five) found it easy; other 11% (two) had no answer. The fact that it was the first time using the system for almost every sample subject (only two had searched on the OCLC regular terminal before), and 72% of them either considered the search procedure not very difficult or easy, suggests that casual users can learn to use the system without facing too many difficulties. When they were asked the question whether they would use the system in the future, six of them (33%) felt they would, eleven (61%) indicated "maybe," only one (6%) said no they would not. The percentage of users that wanted to use the system again is relatively small. However, it shows that there are two groups of potential users. The first group consists of those who are in the fields of Chinese linguistics and literature. Among those who said they would use the system in the future, three are from the CSL group chosen from the department of East Asian Languages and Literatures of The Ohio State University; this is

60% of the entire CSL sample group. In a large research university like The Ohio State University, this group of users can extend to other departments doing China related studies in such fields as art, history, sociology, etc.

The second group of users is the overseas Chinese. Three of the sample users from Hong Kong indicated they would use the system again, while nine from Taiwan and the PRC said they might. As the Chinese student community from Mainland, Taiwan, Hong Kong, and other Chinese-speaking areas of the world is increasing in North American universities, there should be a growing demand in libraries for the OCLC CJK system. It is difficult, however, to extrapolate on the demand for the system by casual users since the sample population in this study is very small. Further studies are needed in this area.

Table 1. Search Pattern

Group	Input Method			Chara. Form		Homo. Reducer		
	PY	WG	TC	CF	CS	Tone	Not Use	Radical
ML	5	0	0	0	5	4	1	0
TW	5	(2)	0	5	0	5	0	0
HK	1	0	2	3	0	1	0	0
CSL	5	(1)	0	3	2	4	1	0
Total	16	0	2	11	7	14	2	0
	89%		11%	61%	39%	87%	13%	

Table 2. Perception A

Group	Tone Homophone Reducers			CJK Input Code Dictionary		
	Impressive	Helpful	Not Helpful	Easy	Hard	Impossible
ML		4				
TW	4	1				
HK		1		1	1	
CSL		4				
Total	4 29%	9 71%	0	1 50%	1 50%	0

Table 3. Perception B

Group	Searching				Use It Again		
	V Diff.	Not V. Diff.	Easy	No Ans.	Certainly	Maybe	No
ML	0	3	2	0	0	5	0
TW	2	1	1	1	0	4	1
HK	0	2	1	0	3	0	0
CSL	1	2	1	1	3	2	0
Total	3 17%	8 44%	5 28%	2 11%	6 33%	11 61%	1 6%

3. Search Results

In the field test, 90 items were searched in both the CJK and roman modes of the OCLC CJK system with three types of search keys: title, personal or corporate author, name/title. Altogether there were 540 records retrieved from the OCLC Online Catalog (OLUC); 270 resulted from searches in CJK mode, the other 270 came from roman mode searches. For the purpose of generating distribution patterns of the kinds of records, each of the 540 records is categorized in one of the following groups:

Record A indicates a bibliographic record; it is an ideal result in which a first attempt in the search produces a record quickly and precisely.

Record B indicates that an initial attempt fails to retrieve a bibliographic record but gives a listing of truncated records from which the title and author of the item can be identified. Record B, although not as ideal as Record A, is still considered an effective one, for there is only one more step required to retrieve the bibliographic record - to key in the number of the item.

Record C presents a collective record from which an author's name or a title can be recognized, but further searching is required in order to retrieve a bibliographic record. This kind of search can be very time-consuming if there are many items under one author's name or having an identical search key; therefore, it is seen as less effective than Record A and B.

Record D is a " yes or no" record, which means search key produces more than fifty entries, and the search can not be continued although "yes" command is keyed in; therefore, it is a failed search.

Record DC appears to be a "yes or no" record, but searching can be continued when "yes" command is inputted. Bibliographic record can be retrieved eventually

after several attempts; the search is very time-consuming.

Record E indicates material not found, thus the search has failed. This usually happens in two situations: first, the item has not been cataloged: second, its record is not sufficiently cataloged. Details about the second case will be discussed in a later section of the this paper.

In brief, among these six types of records, A,B,C, and DC are considered as material found records; D and E are seen as not found.

1) Distribution of Records

Table 4 presents the distribution of these six types of records retrieved with various kinds of search keys. It shows that the most effective way to retrieve a bibliographic record (Record A) was to use Name and Title search key (N/T) in CJK mode: fifty-seven out of a total of 90 records (63%) are Record A. Second most effective was the CJK mode title (TI) search, which produced 54% of Record A (49 out of a total of 90). On the contrary, N/T and TI searches in roman mode produced much lower percentages of Record A, only 7% and 21% respectively, when compared with their CJK counterparts. Similarly, author (AU) search in roman mode also produced lower percentages of Record A relative to CJK mode author search, only 2%.

Table 4 shows also that roman mode searches always produced higher percentages of Record D (Not Found Record) than CJK. For TI search, it is 6% higher (8% versus 2%); for PN or CN search it is 34% higher (36% versus 2%); for N/T search, it is 38% higher (38% versus 0). Apparently, roman search for Chinese material is at a disadvantage. This disadvantage arises from the design of the search keys and the rules for treatment of Chinese names in roman mode searching.

2) Problems in Roman Search Keys

The three types of roman search keys are author (4,3,1), title (3,2,2,1), and author/title (4,4). In an author search, one uses the first four letters of the last name, followed by a comma, then the first one, two, or three letters (depending on how many letters are in the name) of the author's first name, another comma, and finally the first letter of the author's middle name (which is optional). This search key, which can be used easily in searching Western authors, is insufficient for a Chinese author search.

Chinese names usually consist of three characters; the first of which is family name, the last two of which constitute the given name. The rule says, however, that one should treat the two Chinese characters of the given name as one word, and the diacritic should not be considered when inputting the search key. For example, to search a name Ch'eng, Hua-ch'ien, the search key

is "chen,hua,", the character "ch'ien" is excluded and diacritic " ' " and tone are ignored. Naturally, many entries with an identical head will be pulled out:

chen, hua
chen, hua-ai
...
chen, hua-pin
chen, hua-p'in
...
chen, huan
chen, huan-chi
...
chen, huan-tzu
...
chen, huang
chen, huang-
...
ch'en, hua
ch'en, hua-
...
cheng, hua
cheng, hua-
...
ch'eng, hua
ch'eng, hua-erh
...
ch'eng, hua-chi
ch'eng, hua-chin
ch'eng, hua-ching
ch'eng, hua-ch'i
ch'eng, hua-ch'ien

The list can easily fill up to fifty entries before "Ch'eng, Hua-ch'ien is reached, thus the "yes or no" record is produced. The problem here is that the search key, "chen, hua" unsuccessfully limits the search to the desired form, ch'eng, hua-ch'ien. If the third word, "ch'ien," were counted, the situation would surely improve.

In addition to the problem described above, the problem of

homophones (addressed in an earlier part of this study) also comes into play. The romanization hua, for example, with four different tones can represent more than twenty Chinese characters. Even if only five of these twenty odd characters are likely to be used in a name, the list still grows even longer. That is because a single romanization can be used several times to represent different characters, and therefore, different names:

Chen, hua
Chen, hua
Chen, hua
Chen, hua
Chen, hua
Chen, hua-ai
...

Another issue in Chinese names also makes the roman mode N/T search unfavorable. There are about one hundred Chinese surnames, but only ten to twenty of them are commonly used. If there are 10,000 titles, 8,000 of them could be written by people having these twenty names. Divide 8,000 into 20, one gets 400 authors who share the same surname. Roman mode N/T search key (4,4), using only the first four letters of the author's surname, will certainly produce a long list of authors with the same name, not to mention the same roman form may represent different Chinese surnames. In addition, those surnames that are romanized with more than four letters but the first four are identical to the search key's will have little chance to be included in the first fifty entries. Because of all these particularities of the Chinese language and Chinese names, the more words used in a

search key, the more accurate the search for Chinese materials will be. This is also why roman mode title search is more effective (21% of type A, 58% of type B) than any other kind of roman search; it involves four characters.

Table 4. Distribution Of The Six Types Of Records

S.Key	TI		PN & CN		N/T	
TP	CJK	RM	CJK	RM	CJK	RM
A	49(54%)	19(21%)	28(31%)	2 (2%)	57(63%)	6 (7%)
B	26(29%)	52(58%)	44(49%)	19(21%)	25(27%)	24(27%)
C	5 (6%)	7 (8%)	10(11%)	21(23%)	0	13(14%)
DC	3 (3%)	2 (2%)	1 (1%)	12(14%)	0	9(10%)
D	2 (2%)	7 (8%)	2 (2%)	32(36%)	0	34(38%)
E	5 (6%)	3 (3%)	5 (6%)	4 (4%)	8(10%)	4 (4%)
TL	90 100%	90 100%	90 100%	90 100%	90 100%	90 100%

Table 5. Comparison of Found And Not Found Rates in CJK and Roman Searches

FOUND	CJK	ROMAN	Not Found	CJK	ROMAN
A	134 (49%)	27 (10%)	D	4 (2%)	73 (27%)
B	95 (35%)	85 (32%)	E	18 (7%)	11 (4%)
C	15 (5.5%)	41 (15%)	TOTAL	22 (9%)	84 (31%)
DC	4 (1.5%)	33 (12%)			
TOTAL	248 (91%)	186 (69%)			

3) CJK Search Superior to Roman

A comparison of hit rates in both modes shows that CJK search, as a whole, surpasses roman search by 22%; for Record A, CJK surpasses roman by 39% (See Table 5). In addition to Record A,

Record B results are also indicative of effective and relatively quick searches, since the desired material can be identified immediately. Table 5 indicates that CJK mode searches produce 84% (type A 49% + type B 35) of this kind of result, precise and fast, while roman searches produce only 42%, less than half of the CJK. This information suggests that when searching for Chinese materials, CJK searches are favorable and sometimes necessary since CJK search keys allow the computer to match up to 5 characters instead of 3 or 4 letters.

4) Yes or No Records

There are two kinds of "Yes or no" records when the "yes" command is keyed in: Record DC or Record D. Among Record D however, there are two common possibilities. One is that entries are arranged according to year periods; if one knows the period in which the material is published, one can eventually find what is wanted. Because in this field test, qualifiers were not used, this kind of Record D is treated as not found. The other kind of D record is "impossible record." After the "yes" command is inputted, the computer screen shows "request impossible" because the search key produces more than 1,500 records; the search is therefore discontinued. However, there are only three "impossible" records among the total of 77 Record D results (3.9%). It suggests that casual users should be informed about the importance of obtaining the year of publication as basic citation information along with title and author.

5. Record E

Record E is a different kind of not-found record from Record D. When Record E happens a message of "so and so is not in such and such index" will appear on the computer screen. There is a total of 29 Record E results (3.4% of 540 records), 18 from CJK (6.6% of 270 records), 11 from roman search (4% of 270). It was mentioned earlier that Record E happens in two occasions; first when an item has not been cataloged, secondly, when an item is not sufficiently cataloged. The first case is not the concern of this study since all items used in the field test were taken from the OCLC Chinese Online Catalog. In the second case, sample users came across three major types of insufficient cataloging during the field test. These types are described as follows.

A. There is no CJK record established.

1. Record is in roman form, CJK characters have not yet been added; hence, CJK search will produce E record, while roman search can be successful.

2. An item has both Chinese and English titles, but it is cataloged only in English; both the Chinese characters and their romanizations are ignored.

These two are the major causes of the higher percentage of Record E in CJK search.

B. Entries are cataloged in an irregular romanization form.

For example, one would expect to search for the Chinese name 王仲殊, under Wang, chung-shu, or 赵元任, under Chao, y"uen-

jen, the Wades-Giles romanization. However, both of them are entered in the form of Pinyin, for whatever reason, as Wang, zhong-shu and Chao, y"uan-ren. As a result, both personal name searches in roman mode failed.

C. Insufficient added entries:

a. Editor is not traced, thus its name is not in the author index.

b. Only editor of the early volume is traced, others are not. When one searches the later volumes by their editor, record cannot be found.

VIII. Conclusion

This study was conducted to examine whether the OCLC CJK system, so far an exclusive librarian tool, can also serve as a public online Chinese union catalog for Chinese language users.

The results show the following:

1. The two input methods, Pinyin and Ts'ang-chieh enable both the mandarin and non-mandarin speakers to search on the OCLC CJK system;

2. The availability of two forms of Chinese characters seems to allow an individual to retrieve materials in a character form with which he or she is not familiar;

3. The CJK Input Code Dictionary was used by only two subjects; thus, opinions on the dictionary have not been adequately collected and further studies are necessary;

4. Tone homophone reducer seems welcomed by most of the users because it effectively eliminates homophones and streamlines the search process;

5. Because of its unique search keys, CJK search, in general, is more effective in searching Chinese materials than roman search;

6. More than half of the subjects considered searching on OCLC CJK system either not very difficult or easy;

7. Only one subject felt he would not use the system again in the future; 33% of them believed they would and 61% said maybe.

In conclusion, the OCLC CJK system can be beneficial to Chinese language users, as an online Chinese union catalog; its CJK search in many cases is favorable and sometimes necessary. This study shows that potential public users exist, however, for more precise estimate of the public demand of the system further studies are necessary. According to the findings of this study, public users can learn to use the OCLC CJK Workstation without confronting too many difficulties, when necessary instruction is supplied. This study also suggests that materials such as Pinyin and Mandarin Phonetic Symbols Table, or a Chinese dictionary could be helpful for CJK system users for selecting input method, constructing search key, cleaning up confusions in spelling, pronunciation, form of Chinese characters. They can be placed right next to the terminal with The CJK Input Code Dictionary and the user' manual for convenient consulting.

Bibliography

Kim, Jeong Hyun. (1988). "The Search Failure of OCLC CJK350 System: A Pilot Study on User search Pattern of Korean Vernacular Records". A Master's Research Paper submitted to the Kent State University School of Library Science.

Kwei, Chih-ber. (1931). Bibliographical and Administrative Problems Arising from the Incorporation of Chinese Books in American Libraries. China: The Leader Press.

Lee, Thomas H. (1988). The Development of CJK Bibliographic Databases in North America and East Asia. Cataloging and Classification Quarterly Vol.8, (March-April), 111-126.

Liang, Shih-chiu. ed. (1972). A New Practical Chinese-English Dictionary. Taiwan: The Far East Book.

OCLC ed. (1987). CJK350 Input Code Dictionary. Ohio: OCLC.

____ ed. (1986) CJK350 User Manual. Ohio: OCLC

Norman, Jerry. (1988). Chinese. New York: Cambridge University Press.

Tucker, Alan. (1982) "The East Asian Project of the Research Libraries Group". A paper presented at the ASIS' 82 Conference in Columbus, Ohio.

Wang, Andrew H. (1986). " OCLC CJK Automated Library Information Network", A paper presented at the 38th Annual Meeting of Association for Asian Studies.

____ (1988). The OCLC Library Network of Chinese, Japanese and Korean Characters. British Library Occasional Papers;

Ten Chinese Studies. England: The British Library, 210-216.

Wei, Karen T. (1986). Current Status and Future Trends of East Asian Library Automation in North America. Information Technology and Libraries. (June) 140-144.

____ and Sachie Noguchi. (1989). RLIN CJK versus OCLC CJK: The Illinois Experience. Library Resources & Technical Services. Vol. 33, (April) 140-150.

Appendix: A

Search Sheet NO: Name:
Part I. Searching forms:
1. A. CJK Mode Searching

Title:

---

a. Input Methods:
Pinyin (PY) ____
Wade-Giles(WG) ____
Ts'ang-Chieh (TC) ____
b. Script forms:
Full character (CF) ____
Simplified character (CS) ____
c. Input character codes:
ti:____ _____ _____ _____ _____
e. Result: Found ____ (with printout).
Not Found ____ (with printout).

B. Roman Mode Searching

Title:

---

a. Input codes: ____ ____ ____ ____
b. Result: Found ____ (with printout)
Not Found ____ (with printout)

---

2. A. CJK Mode Searching

Personal Name:

---

a. Input Methods:
Pinyin (PY) ____
Wade-Giles(WG) ____
Ts'ang-Chieh (TC) ____
b. Script form:
Full character (CF) ____
Simplified character (CS) ____
c. Input character codes:
pn:_____ _____ _____ _____
e. Result: Found ____ (with printout).
Not Found ____ (with printout).

B. Roman Mode Searching

Personal Name:

a. Input codes: ____ ____ ____
b. Result: Found ____ (with printout)
Not Found ____ (with printout)

2. A. CJK Mode Searching

Corporate Name:

a. Input Methods:
Pinyin (PY) ____
Wade-Giles(WG) ____
Ts'ang-chieh (TC) ____
b. Script form:
Full character (CF) ____
Simplified character (CS) ____
c. Input character codes:
cn:_____ _____ _____ _____
e. Result: Found ____ (with printout).
Not Found ____ (with printout).

B. Roman Mode Searching

Corporate Name:

a. Input codes: =____ ____ ____
b. Result: Found ____(with printout)
Not Found _____ (with printout)

3. A. CJK Mode Searching

Name/Title:

a. Input Methods:
Pinyin (PY) ____
Wade-Giles(WG) ____
Ts'ang-chieh (TC) ____
b. Script form:
Full character (CF) ____
Simplified character (CS) ____
c. Input character codes:
nt:_____ , _____ _____ _____ _____
e. Result: Found ____ (with printout).
Not Found ____ (with printout).

B. Roman Mode Searching

Name/Title:

a. Input codes: _____ _____
b. Result: Found _____(with printout)
Not Found _____ (with printout)

```
 Screen 1 of 3
▸OSU - FOR OTHER HOLDINGS, ENTER .dh DEPRESS DISPLAY RECD SEND
 OCLC: 16670113      Rec stat: n Entrd: 870914          Used: 890602
▸Type: a Bib lvl: n Govt pub:   Lang:  chi Source: d Illus:
 Repr:    Enc lvl: I Conf pub: 0 Ctry:  ch  Dat tp: s M/F/B: 10
 Indx: 0 Mod rec: r Festschr: 0 Cont:
 Desc: a Int lvl:   Dates: 1967,
 ▸ 1 010
 ▸ 2 040      OSU *c OSU
 ▸ 3 090      Z3101 *b .S4 1967
 ▸ 4 090      *b
 ▸ 5 049      OSUU
 ▸ 6 100 10   Shao, I-ch"en, *d 1810-1861.
 ▸ 7 100 10   邵懿辰, *d 1810-1861.

 Screen 2 of 3
 ▸ 8 245 10   Tseng ting ssu k"u chien ming mu lu piao chu / *c
Shao I-ch"en chuan ; Sun I-jang teng ts"an chiao ; Shao Chang hs"u lu, Shao
Yu-ch"eng ch"ung pien.
 ▸ 9 245 10   增訂四庫簡明目錄標注 / *c 邵
懿辰撰 ; 孫詒讓等參校 ; 邵章續錄, 邵友誠重
編.
 ▸10 250      Tsai pan.
 ▸11 250      再版.
 ▸12 260 0    T'ai-pei shih : *b Shih chieh shu ch"u, *c min kuo
56 [1967]
 ▸13 260 0    臺北市 : *b 世界書局, *c
民國56 [1967]
 ▸14 300      2 v. (2, 11, 1038 p.) ; *c 19 cm.
 ▸15 490 0    Chung-kuo hs"ueh shu ming chu. Ti 3 chi
 ▸16 490 0    中國學術名著. 第3輯

 Screen 3 of 3
 ▸17 490 0    Chung-kuo mu lu hs"ueh ming chu. Ti 2 chi : *v ti 5
ts"e
 ▸18 490 0    中國目錄學名著. 第2集
; *v 第5冊
 ▸19 500      Running title: Tseng ting ssu k"u chien mu piao
chu.
 ▸20 500      Running title: 增訂四庫簡目標注.
 ▸21 651  0   China *x Bibliography.
 ▸22 650  0   Chinese literature *x Bibliography.
 ▸23 700 10   Shao, Chang, *d chin shih 1903.
 ▸24 700 10   邵章, *d chin shih 1903.
 ▸25 700 10   Shao, Yu-ch"eng.
 ▸26 700 10   邵友誠.
```

Figure 1. Sample of OCLC CJK Record

Tseng ting ssu k'u chien ming mu lu piao chu

Z 3101
S4
1967

Shao, I-ch'en, 1810-1861.
(Tseng ting ssu k'u chien ming mu lu piao chu)
增訂四庫簡明目錄標注 / 邵懿辰撰 ; 孫詒讓等參校 ; 邵章續錄, 邵友誠重編. — 再版. — 臺北市 : 世界書局, 民國56 [1967]
2 v. (2, 11, 1038 p.) ; 19 cm. — (中國學術名著. 第3輯) (中國目錄學名著. 第2集 ; 第5冊)
Running title: 增訂四庫簡目標注.

1. China--Bibliography. 2. Chinese literature--Bibliography. I. Shao, Chang, chin shih 1903. II. Shao, Yu-ch'eng. III. Title

OSU arc870914

Figure 2. Sample of OCLC CJK Card

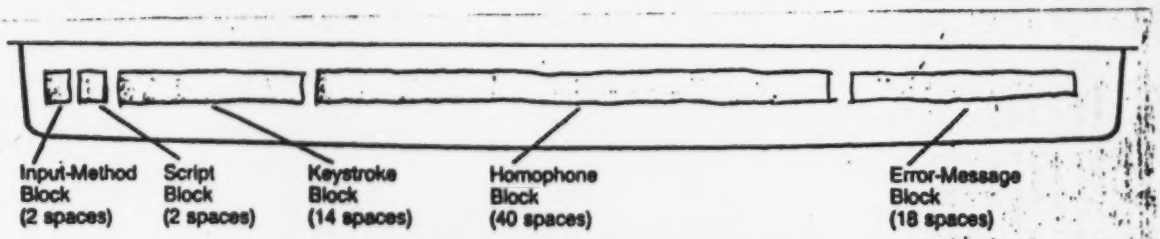

Figure 3.1 Message Blocks in the Input Status Line

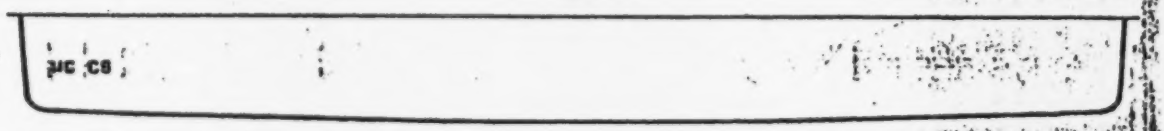

Figure 3.2 Input Method and Script Selected

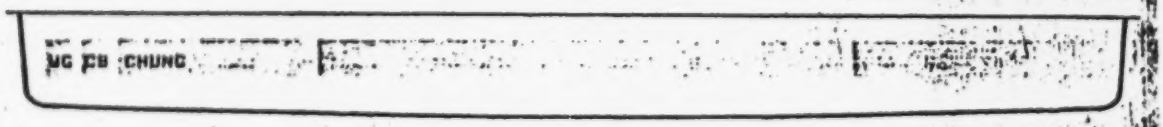

Figure 3.3 Input Code Typed

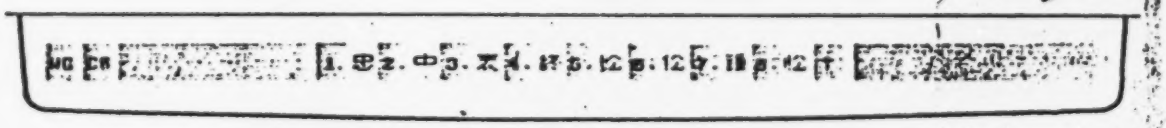

Figure 3.4 Matching Characters Displayed in the Homophone Block

Figure 3. The CJK Message Block in the Status Line

BEST COPY AVAILABLE

Graphic Elements Assigned to English-Letter Keys for Tsang-chieh Entry

English Letter	Primary Element	Secondary Element	English Letter	Primary Element	Secondary Element
A	日	田	M	一	工厂
B	月	冂冂⺆多	N	弓	乙乛
C	金	钅ソハ儿	O	人	亻𠂉くくㇵ
D	木	才艹	P	心	忄小匕匕七勹七十
E	水	氵又	Q	手	扌扌扌牛
F	火	小小灬灬	R	口	
G	土	土	S	尸	コ匚冂尸
H	竹	𥫱丿𠂉	T	廿	廿卅丗艹
I	戈	丶厶𠂉	U	山	凵乚丿丨屮
J	十	𠂇	V	女	𠃌く𠂢
K	大	又𠂇	W	田	口口
L	中	丨丿⺶	Y	卜	卜ユ丷冫ネニ

Figure 4. Table of Graphic Elements Assigned to English Letter for Ts'ang-chieh Entry

BEST COPY AVAILABLE

一、字 母 表

字母	Aa	Bb	Cc	Dd	Ee	Ff	Gg
名称	ㄚ	ㄅㄝ	ㄘㄝ	ㄉㄝ	ㄜ	ㄝㄈ	ㄍㄝ
	Hh	Ii	Jj	Kk	Ll	Mm	Nn
	ㄏㄚ	ㄧ	ㄐㄧㄝ	ㄎㄝ	ㄝㄌ	ㄝㄇ	ㄋㄝ
	Oo	Pp	Qq	Rr	Ss	Tt	
	ㄛ	ㄆㄝ	ㄑㄧㄡ	ㄚㄦ	ㄝㄙ	ㄊㄝ	
	Uu	Vv	Ww	Xx	Yy	Zz	
	ㄨ	ㄪㄝ	ㄨㄚ	ㄒㄧ	ㄧㄚ	ㄗㄝ	

v只用来拼写外来语、少数民族语言和方言。

字母的手写体依照拉丁字母的一般书写习惯。

二、声 母 表

b	p	m	f	d	t	n	l
ㄅ玻	ㄆ坡	ㄇ摸	ㄈ佛	ㄉ得	ㄊ特	ㄋ讷	ㄌ勒
g	k	h		j	q	x	
ㄍ哥	ㄎ科	ㄏ喝		ㄐ基	ㄑ欺	ㄒ希	
zh	ch	sh	r	z	c	s	
ㄓ知	ㄔ蚩	ㄕ诗	ㄖ日	ㄗ资	ㄘ雌	ㄙ思	

在给汉字注音的时候，为了使拼式简短，zh ch sh可以省作ẑ ĉ ŝ。

三、韵 母 表

	i ㄧ 衣	u ㄨ 乌	ü ㄩ 迂
a ㄚ 啊	ia ㄧㄚ 呀	ua ㄨㄚ 蛙	
o ㄛ 喔		uo ㄨㄛ 窝	
e ㄜ 鹅	ie ㄧㄝ 耶		üe ㄩㄝ 约
ai ㄞ 哀		uai ㄨㄞ 歪	
ei ㄟ 欸		uei ㄨㄟ 威	
ao ㄠ 熬	iao ㄧㄠ 腰		
ou ㄡ 欧	iou ㄧㄡ 忧		
an ㄢ 安	ian ㄧㄢ 烟	uan ㄨㄢ 弯	üan ㄩㄢ 冤
en ㄣ 恩	in ㄧㄣ 因	uen ㄨㄣ 温	ün ㄩㄣ 晕
ang ㄤ 昂	iang ㄧㄤ 央	uang ㄨㄤ 汪	
eng ㄥ 亨的韵母	ing ㄧㄥ 英	ueng ㄨㄥ 翁	
ong (ㄨㄥ) 轰的韵母	iong ㄩㄥ 雍		

附錄 國語注音符號與各式羅馬拼音對照表

一、聲 母

國語注音符號 Mandarin Phonetic Symbols	威妥瑪式 Wade System	國語羅馬字 Kwoyeu Romatzyh	耶魯大學式 Yale System	國語注音符號 Mandarin Phonetic Symbols	威妥瑪式 Wade System	國語羅馬字 Kwoyeu Romatzyh	耶魯大學式 Yale System
ㄅ	p	b	b	ㄏ	h	h	h
ㄆ	p'	p	p	ㄐ	ch(i)	j(i)	j(i)
ㄇ	m	m	m	ㄑ	ch'(i)	ch(i)	ch(i)
ㄈ	f	f	f	(ㄗ)		(gu)	
(ㄖ)		(v)		ㄒ	hs	sh(i)	s(i)
ㄉ	t	d	d	ㄓ	ch	j	j
ㄊ	t'	t	t	ㄔ	ch'	ch	ch
ㄋ	n	n	n	ㄕ	sh	sh	sh
ㄌ	l	l	l	ㄖ	j	r	r
ㄍ	k	g	g	ㄗ	tz, ts-	tz	ds
ㄎ	k'	k	k	ㄘ	tz', ts'-	ts	ts
(ㄨ)		(ng)		ㄙ	ss, ss, s-	s	s

二、韻 母

國語注音符號 Mandarin Phonetic Symbols	威妥瑪式 Wade System	國語羅馬字 Kwoyeu Romatzyh	耶魯大學式 Yale System	國語注音符號 Mandarin Phonetic Symbols	威妥瑪式 Wade System	國語羅馬字 Kwoyeu Romatzyh	耶魯大學式 Yale System
(ㄭ)	-i, -ih	y	i, r	ㄧㄡ	yu, -iu	iou	you
ㄚ	a	a	a	ㄧㄢ	yan, -ien	ian	yan
ㄛ	o	o	o(wo)*	ㄧㄣ	yin, -in	in	yin, -in
ㄜ	e(o)	e	e	ㄧㄤ	yang, -iang	iang	yang
ㄝ		ê	ê	ㄧㄥ	ying, -ing	ing	ying, -ing
ㄞ	ai	ai	ai				
ㄟ	ei	ei	ei	ㄨ	wu, -u	u	wu, -u
ㄠ	ao	au	au	ㄨㄚ	wa, -ua	ua	wa
ㄡ	ou	ou	ou	ㄨㄛ	wo, -uo	uo	wo
ㄢ	an	an	an	ㄨㄞ	wai, -uai	uai	wai
ㄣ	en	en	en	ㄨㄟ	wei, -ui(-uei)	uei	wei
ㄤ	ang	ang	ang	ㄨㄢ	wan, -uan	uan	wan
ㄥ	eng	eng	eng	ㄨㄣ	wen, -un	uen	wen
ㄦ	erh	er	er	ㄨㄤ	wang, -uang	uang	wang
				ㄨㄥ	weng, -ung	ueng, -ung	weng, -ung
ㄧ	i	i	yi, -i				
ㄧㄚ	ya, -ia	ia	ya	ㄩ	yu, -ü	iu	yu
ㄧㄛ		io		ㄩㄝ	yueh, -üeh	iue	yue
ㄧㄝ	yeh, -ieh	ie	ye	ㄩㄢ	yüan, -üan	iuan	yuan
ㄧㄞ	yai	iai	yai	ㄩㄣ	yün, -ün	iun	yun
ㄧㄠ	yao, -iao	iau	yau	ㄩㄥ	yung, -iung	iung	yung

註：*wo* 在ㄅㄆㄇㄈ之後用。

Figure 5. Pinyin, Mandarian Phonetic Symbols, and Wade-Giles Tables

BEST COPY AVAILABLE